

Informed consent for: “The ethos and effects of data-sharing rules: Examining the history of the ‘Bermuda principles’ and their effects on 21st century science”

University of Adelaide

Duke University

Researchers at the University of Adelaide, Australia, and the IGSP Center for Genome Ethics, Law & Policy, Duke University, are engaged in research on the **Bermuda Principles** for sharing DNA sequence data from high-volume sequencing centers. You have been selected for an interview because we believe that the recollections you may have of your experiences with the International Strategy Meetings for Human Genome Sequencing (1996-1998) will be interesting and helpful for our project.

We expect that interviews will last from 30 minutes to much longer, but you may stop your interview at any time. Your participation is strictly voluntary, and you do not have to answer every question asked.

Your interview is being recorded and we may take written notes during the interview. After your interview, we may prepare a typed transcript of the interview. If we prepare a transcript, you will have an opportunity to review it and to make deletions and corrections.

Unless you indicate otherwise, the *information* that you provide in this interview will be “on the record”—that is, it can be attributed to you in the various articles and chapters that we plan to write, and thus could become public through these channels. If, however, at some point in the interview you want to provide us with information that might be useful for us to know, but which you do not want to have attributed to you, you should tell us that you wish to go “off the record” and we will stop the recording. We will, however, take notes for our own use. When you are ready to go back “on the record,” we will resume recording. Anything you say while “off the record” will not be on the audio recording and therefore will not appear in the transcript.

All *materials* from your interview (audio recording; transcript; interviewer's notes) will be available only to members of the research team affiliated with this project, unless you consent to their wider use, as described in the paragraph below. The digital materials will be maintained in a secure, HIPPA-compliant drive at Duke University. The paper materials will be stored in a locked cabinet.

In addition to the scholarly articles and chapters that we plan to write, we also hope to create a resource for other scholars and members of the public. We plan to post some of our research data to online digital archives. While we will use your “on the record” comments to inform and write our articles, we will not post your interview transcript or audio recording online unless you give us permission to do so, in a separate agreement. At the time we send your transcript to you for review, we will also provide a consent form asking your permission to post your interview transcript and/or audio recording online. The form will provide you with different options for how, when, and with whom the materials may be shared. You will, of course, also have the option not to share the materials beyond the Duke and Adelaide researchers.

One risk of this study is that you may voluntarily disclose identifiable information that later could be requested for legal proceedings, or otherwise be used against you. Please take this into consideration when you are speaking. There may be other risks associated with your “on the record” views being made publicly available, such as having your views mischaracterized or misunderstood.

The main benefit of participating in this study is ensuring that your side of the story is properly portrayed in this history of the Bermuda Principles, which have become a model for open and collaborative research in genomics and other fields.

To help us protect the privacy of those parts of your interview that are not public, we have obtained a Certificate of Confidentiality from the U.S. National Institutes of Health. With this Certificate, we investigators cannot be forced to disclose information that may identify you, even by a court subpoena, in any U.S. federal, state, or local civil, criminal, administrative, legislative, or other proceedings. We researchers can use the Certificate to resist any demands for information that would identify you.

The Certificate cannot be used, however, to resist a demand for information from personnel of the United States Government that is used for auditing or evaluation of federally funded projects or for information that must be disclosed in order to meet the requirements of the federal Food and Drug Administration (FDA).

A Certificate of Confidentiality does not prevent you or a member of your family from voluntarily releasing information about yourself or your involvement in this research. If an insurer, employer, or other person or institution obtains your written consent to receive research information, the researchers may not use the Certificate to withhold that information.

Signature  _____

Printed Name C. Thomas Caskey _____

Date 3/2/12 _____

If you have read this form in its entirety and agree to the interview and its terms, please sign and date above.

Contact information:

Rachel Ankeny, Ph.D. (University of Adelaide)

rachel.ankeney@adelaide.edu.au

+61-8-8303-5570

Kathryn Maxson, B.S. (Duke University)

kat.maxson@duke.edu

(919) 668-0791

Robert Cook-Deegan, MD (Duke University)

bob.cd@duke.edu

(919) 668-0790

*If you have any questions about your rights as a research subject, you may contact the **Duke University Institutional Review Board** at 919-684-3030 or ors-info@duke.edu.*

PLEASE FILL OUT AND RETURN THIS FORM TO: Center for Public Genomics, Duke University; c/o Susan Brooks; Center for Genome Ethics, Law, and Policy; 304 Research Drive, Box 90141; Durham, NC, 27708. **OR:** You may fax it to us at (U.S.) 1-919-668-0799.

Interviewee Information. Please list an address where we can contact you.

Full name: C. Th. Caskey Date of interview: 3/2/12
Current institutional affiliation: Baylor College of Medicine
Street Address: One BAYLOR PLAZA, BCM 225, HOUSTON TX 77030
Phone: 713-798-3186 Email address: tcaskey@bcm.edu

Interviewer Information.

Full name(s): _____
Affiliations(s): _____

I, the undersigned, have read the above, and I **AGREE** to release my interview materials, subject to any restrictions listed below:

(A) I place **no restrictions** on my interview materials.

OR

(B) My interview materials may be reviewed, used, and quoted by the researchers affiliated with the Center for Public Genomics, Duke University; *and in addition* (check all that apply):

Researchers unaffiliated with the Center for Public Genomics may **read** the interview transcript and any related documents only after obtaining my permission.

Researchers unaffiliated with the Center for Public Genomics may **quote** from the interview only after obtaining my permission.

ok Researchers unaffiliated with the Center for Public Genomics **DO NOT HAVE** my permission to **read or quote** from the interview.

Posting interview materials to public digital archives: In spite of any restrictions listed above, I give permission for my interview materials to be made publicly available on the Internet by deposit in an institutionally affiliated archive:

1 year from the date of this form

5 years from the date of this form

10 years from the date of this form

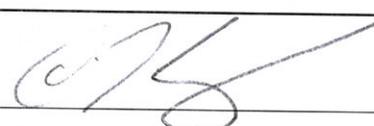
25 years from the date of this form

After my death

Other: _____ (please specify a date or condition)

NEVER: MAY NOT BE DEPOSITED IN A PUBLIC ARCHIVE

Please specify any further restrictions in the space below:

Signature: 

Date: 3/23/12

Interviewee: C. Thomas Caskey

Date, location, method: 02 March 2012, Durham, NC, by phone

Interviewers: Kathryn Maxson, Robert Cook-Deegan

KM: Okay, so we have [BCD] and [KM] here with [TCaskey] from Baylor. And we've explained the informed consent so we're going to move on from that. And [TCaskey], we have you as being at the first Bermuda meeting in 1996, and that was February 27 and 28, 1996. That was the only meeting we have you at, from our records. And you've just told us that you were very sick with a fever during that time period. But do you by any chance remember in what capacity you were invited to that meeting and whom you were representing, and why you think you were there?

TCaskey: Well, I'm not sure I recall exactly how I got on the list but we were one of the very first institutions to get the genome grant. I was in on the original genome planning committee and was head of the genetics grants section, which took all of the grants at that time. It could have been just the collective involvement in the genome project that got me an invitation. The other thing that was ongoing at that time was we were making a significant number of disease gene discoveries using sequencing in family studies. So the laboratory was very, very active and productive during that time. The department was extremely active and productive during that time, too. So all of that helps.

KM: And this was at Baylor?

TCaskey: This was at Baylor, yes.

KM: So what was your understanding in advance of what was going to be discussed at this meeting in 1996, if you can remember?

TCaskey: Well, it was going to be developing a policy of openness of sharing sequence information. That was really what was on the table.

KM: And what about scientifically? What about some of the scientific issues that you saw were planning on being discussed at the time?

TCaskey: We were so engaged in the science at that point things were moving along pretty well. I don't recall any kind of significant breakthroughs in science that were discussed. We already had PCR, we had the large BAC cloning, we had automated sequencing devices, we had informatics, we were making discoveries. The project was underway.

BCD: And [TCaskey], do you remember where exactly...Baylor was one of the original sequencing centers that had been funded by NIH...well, at that point it was NCHGR.

KM: NCHGR, right.

Interviewee: C. Thomas Caskey

Date, location, method: 02 March 2012, Durham, NC, by phone

Interviewers: Kathryn Maxson, Robert Cook-Deegan

BCD: So if you could just take us back down memory lane here.

TCaskey: Yeah, there's one more element to it, and this might have been the reason that I was...because I left Baylor in '94 to become Senior Vice President for Research at Merck. I was running the West Point site. And we were the funders of the EST project, which was an open-access database to ESTs. And that might have been the true reason that I was invited to that meeting, now that I reflect on it.

KM: Well you know, under your name for the invitee list for that meeting, you're listed as being from Merck and not from Baylor. So that would make sense.

TCaskey: Yeah, and I think we were the sole and major funder of the EST project.

KM: Okay.

TCaskey: And it was run predominantly out of Wash U. That's probably the reason I was invited to the conference.

KM: Okay. So would you mind, just for the tape, describing a little bit more about that project? What the background was and the purpose?

TCaskey: Well we knew that the technology was not up to doing genome sequencing at that point. It was too expensive and too labor intensive, and it really hadn't come to the point where we had the strategy on how to do whole genomes very efficiently. But we knew that the action was in genes; and therefore, our interest from Merck, and I think it may have been several other pharmaceutical companies that joined us on this project, was that we would have open access to sequence information on genes. Because genes, of course, are the targets we used for the development of drugs. So we had a vested interest in making sure that this information was open access for all to use in promoting the development of new drugs. In other words, the tool would be available to you without restriction. So that was probably our unspoken objective in funding an open-access database: to get knowledge of a gene's structure. And of course, that was doable by the technology at the time. You could copy the transcripts very easily, get the fragments and run them through sequencing. Of course that task was, what...about a hundred-fold less than the task of trying to sequence the genome, and our feeling was that the genes, it was all about genes, that was the important and critical issue to remain unencumbered by patents at the time. That's my recollection of it.

KM: Right, right. [BCD], do you have anything to add or any questions about that?

BCD: No. And what would probably be useful is, just kind of give us a picture of what it felt like to be in this space at this time and as things were being poised to move on the sequencing.

Interviewee: C. Thomas Caskey

Date, location, method: 02 March 2012, Durham, NC, by phone

Interviewers: Kathryn Maxson, Robert Cook-Deegan

TCaskey: Well first of all, I thought it was an exciting time for identifying new targets for drug development. We were very enthusiastic about this strategy. And bioinformatics was starting to become more, I guess, more efficient at being able to identify genes with functions. For example, at that time I think 70 percent of all drugs were targeted against G protein-coupled receptors. And if you had transcript information it was possible bioinformatically to identify a G protein-coupled receptor transcript without any difficulty at all. There were algorithms you could use to identify them. So the most rich target source for drug development at that time were G protein-coupled receptors; therefore, getting transcript sequences enabled you to mine that database to identify the plethora of G protein-coupled receptors that were out there.

Now of course what we found very quickly was that there were a large number of receptors without known ligands or without known functions. You couldn't predict the functionality of the G protein couple receptor on sequence. That was not possible to do at that time. We didn't have enough sequence information to do comparisons. But what it did do was develop a major program within Merck of developing full-length coding sequences on G protein-coupled receptors and de-orphanizing those receptors. And Jim Liu was head of that program. He had a small team of about four people. He was very, very good working with the chemists and being able to identify families of compounds that we knew were ligands for known receptors. Thus we could use the transcripts and cell-based assays to express a receptor, and then using cocktails of ligands we could identify, maybe not the natural ligands, but we could identify a drug-like ligand that signaled that receptor. And so we were very successful at that. We filed numerous patents based upon that discovery. Merck has a huge inventory of matching the G protein-coupled receptor to the ligand molecule and to the place in which the receptor was expressed. So we were able to use the information in the development of new therapeutics and we did it derivative of the open access EST project.

KM: That's really interesting. Especially in light of your work at Merck—I've got you as a session chair. I've just been looking through the 1996 program for this meeting. It has you listed as a session chair for large-scale sequencing, discussing the current status of new approaches to sequencing, directed versus shotgun sequencing strategies, cost efficiency, throughput and data handling, quantity versus quality. And some of the folks that are listed as being in that session are Craig Venter, Lee Hood, Rick Wilson, Jane Rogers, André Rosenthal and several others. Now, were you able to chair that session? Or were you unable to because you were sick? And if so, what do you remember from it?

Interviewer: I was able to chair it and I know why they invited me to do that. We were...I think our laboratory was the first to do a full genome sequence on a disease gene.

Interviewee: C. Thomas Caskey

Date, location, method: 02 March 2012, Durham, NC, by phone

Interviewers: Kathryn Maxson, Robert Cook-Deegan

And it was the Lesch Nyhan gene. We did that in collaboration with an automated sequencing device that was being developed at EMBL. The gene was only 50,000 base pairs. As I recall it took us almost six months. But the way we did it was by fragmentation and shotgun sequencing, and then completing the sequence by what we called gap mapping and closure of the gaps. So almost all of the strategy on shotgun sequencing that was being discussed at that session had been prototyped and done on a disease gene, the HPRT gene.

Now of course, the whole hope was doing that on genome sequencing, and we focused in on a gene that we sequenced, but to do the whole genome was a hell of a lot more challenging undertaking. And a lot of the debate that took place at that time...there were two major issues I remember from the meeting...and that is, one, do you organize the clones in an orderly manner and sequence, or do you just do random sequencing? And of course NIH at that point was into organizing the clones, ordering the clones in BACs and cosmids, and then linearly carrying out sequencing on those overlapping clones.

Craig Venter had a totally different strategy, and that was—shotgun the whole genome and sequence. And there was a huge amount of debate that took place at that time on whether the computers had the capacity to be able to handle the shotgun sequence and assemble it. And Craig was absolutely confident that that could be done, and of course he was right. And there was a lot of doubt at the meeting as to whether the computers could handle the high repeats that existed within the genome. Those were big concerns that existed. So do you do it as true shotgun or do you organize the clones and linearly sequence? Because they felt that the informatics challenge of sequencing organized linear clones would be easier than it would to assemble the shotguns. I remember all that. And that was just greatly debated. People had different opinions on it. And Craig ended up being right on the strategy.

BCD: Had you done that work on HPRT at Baylor or at Merck or at some combination?

TCaskey: What we did, we had cloned the gene and I had been involved in working with the EMBL on another project. And they came to me and said, “We’ve got a prototype automated sequencing machine, why don’t you bring this project to EMBL and let’s do the project there?” So we had, I think, a NATO travel grant, and one of my MD PhD students, Al Edwards, did all the traveling back and forth between the two groups. And so it was a collaboration between EMBL and Baylor to do it. We didn’t do it on the ABI instruments. We did it on what was a...I think it was a Pharmacia prototype. But it was very similar technology, very similar technology.

BCD: This was the Ansorge machine?

Interviewee: C. Thomas Caskey

Date, location, method: 02 March 2012, Durham, NC, by phone

Interviewers: Kathryn Maxson, Robert Cook-Deegan

TCaskey: Right, that's exactly right. It was Ansorge. But just think about that. I mean, we crank out a whole genome in 24 hours now, and it took us six months to assemble a 50,000-base pair gene. We weren't too efficient at that time.

KM: So this was 1996, before the whole Celera debacle. What did Craig seem like? Was any of this debate going on right in your session?

TCaskey: Well I think in the early days you had confidence on both sides, both NIH and confidence on Craig's side, that they had the preferred way of doing it. And of course, NIH was going to be distributive in the way they did it. They would involve many different laboratories and getting the clones, and there was a chromosome 15 group and a chromosome 18 group and 21 group. So they were responsible for pulling those clones out and then arranging BACs, at that time, in a linear order to feed into the sequencing machines. There was a big national effort to get the overlapping clones. And it was distributed. I mean—many different laboratories had received grants to do that. And of course what Craig did, was industrialize the project and put all of the activity in one location with the power of the sequencing in that one location and the power of the informatics in that same location. So he had a totally different way of going about it than the way NIH was doing it. He was doing it industrial style and NIH was continuing to do it as a classic research project style.

KM: Right. So do you have any recollection of the other sessions that you attended?

TCaskey: Well I can...because I was very interested in the IP issues on the genome I attended all of those sessions. And I guess my feeling was at the time, it was a little bit different feeling than an attitude. Merck wanted open access to everything. And that's one of the reasons we funded the EST project. NIH was not funding that project, by the way. It was more focused on doing the sequencing on the nuclear DNA as opposed to the expressed sequences. But we came in and said, look, it's all about the genes, let's sequence the genes. And the way to do that was through ESTs. So from the corporate point of view that was a very good effort to keep the knowledge in open access for all to work on.

Now, I had just come out of the experience of making a significant number of disease gene discoveries, which we patented. And those patents held and the medical school did extremely well on those patents. And one of the most important patents, I guess, that we had—we broke the mechanism of the triplet repeat diseases. I also appreciated that this could be used as a personal identification system. So we developed the STR personal identification system, which is now the worldwide standard for personal identification. We patented that system. That patent was licensed by ABI and licensed by Promega. And my recollection of it is they paid something like \$30,000 or \$40,000 for licensing the

Interviewee: C. Thomas Caskey

Date, location, method: 02 March 2012, Durham, NC, by phone

Interviewers: Kathryn Maxson, Robert Cook-Deegan

patent but the royalties were extremely rich on it. And of course this was the singular method that was accepted worldwide. And so the amount of money that came in on that patent was very large. I gave up all my rights to the MD PhD program, so I didn't profit on it. My wife always asks me, why did I do that? But the income from that patent was averaging approximately \$1 million a year to support the MD PhD program. So I was very much in favor of protecting discoveries. Thus I was of two minds. My previous experience said, have a good time making discoveries and protect what you have and try to commercialize it. With Merck it was, for Lord's sake, don't restrict us on doing anything to develop new drugs. We want open access. And that was I think the predominant attitude at that point, so I have lived in both worlds.

BCD: Do you know if this was the STR patent that was the basis of the ABI-Promega suit that was just settled a couple months ago?

TCaskey: We're trying to figure that out, and I thought it was. The lawyers are looking into it, but they're not sure it's our patent. They think it might be some other patent that was in conflict. I can tell you how our patent worked. We licensed it to ABI for fluorescent detection. We licensed it to Promega for radioisotopic detection. And so each of them had a go at it. Now it turns out fluorescents won the day with the machines, with the ABI machines. But I think that Promega—it's possible Promega was making the kit that enabled the fluorescents. I think that might have been going on, but I don't know for sure. It's something that needs to be looked into. But the lawyers at Baylor are in fact looking at it. I don't have an answer.

KM: So you're saying you have these two experiences that spoke to opposing camps in terms of their views on patenting and open access. And a lot of the folks we've talked to, and also in the literature about the Bermuda Principles, indicate that gene patenting was an issue that was one of the causes of this meeting and one of the things anticipated for discussion. What was your sense of that?

TCaskey: Yeah, all right, so I'll tell you a personal experience with that and you can see how this can become a very frustrating situation. We were attempting to discover the Duchenne muscular dystrophy gene, when it in fact was discovered by Lou Kunkel. And Lou patented that gene. It's a large gene and we were very much into the technology of PCR and how you apply PCR to diagnostics. So we developed a way to scan this two million-base pair gene for mutation in a single assay. It was called multiplex amplification. So we delineated where the gene was vulnerable from deletions. We put amplification units down on that and we ran them all in one cocktail, and we had a very, very slick and accurate way of diagnosing Duchenne, of prenatally diagnosing it and finding carriers. It was terrific. So we were doing a lot of diagnosis out of the Baylor diagnostic

Interviewee: C. Thomas Caskey

Date, location, method: 02 March 2012, Durham, NC, by phone

Interviewers: Kathryn Maxson, Robert Cook-Deegan

laboratory, the deoxyribonucleic acid diagnostic laboratory, when one day we get a notice that we have to stop rendering the diagnosis on Duchenne. We have to stop delivering patient services.

BCD: Just from Athena?

TCaskey: Exactly. Athena had licensed the Duchenne muscular dystrophy gene diagnostic rights from Lou, and regardless of the technology we had added that enabled the ability to make the diagnosis, because the patent didn't allow that, we were stopped. And I was always furious over that because it was something that was good for the patient, it was working well, it was highly accurate, it was better than anything that was out there at the time, and we were stopped cold. Our lawyers told us cease and desist. Never was very happy about that. But multiplex amplification got patented. Everybody that amplified more than two things had to pay the medical school for that use.

KM: So this was an issue that you had a pretty good idea was going to be discussed when you went to Bermuda and met with these folks.

TCaskey: Oh, yeah. Oh God, yes. I mean, all these events that I'm telling you about were Baylor College of Medicine discoveries and patents out of the medical college. And we were very rapid to apply these technologies to clinical care.

KM: Okay. So just moving forward now to the end of that meeting in 1996, were you in the room when these so-called Bermuda Principles were discussed? It was the last session in the 1996 meeting. Do you remember that?

TCaskey: Yep, I was there, but I don't have a great deal of recollection over the issues. I'm trying to remember who was helping harness this group, and I think it was the director of the Wellcome Trust at that point. Oh God, what was his name ...

KM: Michael Morgan?

TCaskey: Michael Morgan. I think Michael Morgan played a large role in trying to get consensus. And Francis was involved in that. There were a number of other people. It's a vague recollection. But I'd put Michael out front trying to get this thing to come to a consensus opinion.

KM: So just to jog your memory, at that session, John Sulston and Bob Waterston were chairing. And there was a whiteboard where John wrote some version of what eventually became the Bermuda Principles. Is this jogging your memory at all?

TCaskey: I'll tell you, I'm not going to have good recollection of that. I do not remember ... I don't remember that, I'm sorry.

Interviewee: C. Thomas Caskey

Date, location, method: 02 March 2012, Durham, NC, by phone

Interviewers: Kathryn Maxson, Robert Cook-Deegan

KM: Well, that's okay. [BCD], did you want to jump in at all here?

BCD: So the general context for having this meeting was that NCHGR had just let the first of the contracts to scale up the sequencing. And from many of our other interviews it sounds like in part it was trying to do what you would need to allocate the tasks and get things organized and coordinated. And it looks from the agenda like a lot of the discussion was about who can actually do the mapping, who can actually do the sequencing at what level of accuracy, and how are you going to keep track of things. And out of that came this idea of data sharing principles. And can you walk us through your recollection of getting from A to B?

TCaskey: Yeah, I do remember this. I remember that a number of the academic institutions had concern about this because they wanted to make discoveries. And so the way in which you get the information out if you produce that data out of your program, there is an opportunity for publication and receiving recognition for that discovery. And that's classically the way science works. You advance by making a discovery and being individually associated with that project. So this was a mind-changer in the sense that this was going to be kumbaya, everything going into a central reservoir of knowledge and content. And it was very difficult to understand how an individual investigator could advance academically in their career without being able to get credit for doing something. Now it was very clear that NIH held the purse strings and they were going to dictate this policy. I don't think there was any ambiguity about that. If you want to take the money, then this is the way it's going to be. But the academics were still very much thinking about, how can I be individually recognized for discovery and advance careers the way you do it in traditional academic pathway.

Of course, Craig was operating totally differently. He was industrial, where you pool the resources and get a single product out, and the company benefits from the knowledge, not the individuals within the company. So he had a totally different model of how to do this. And he didn't have any I guess holdback within his company on how to do this. But the academics, once they started parsing the work, there was concern about that. I remember those discussions very much.

And so that was one of the reasons that the time delay was set in on release of information. It was done under the guise that, well, you need to have, oh, six months to make sure the sequence is right and then you can release it. But in fact I think what was behind it all was, we'd like to make some discoveries and have the first crack at writing papers related to the work that we put in. Now that's an opinion. But sitting around the bar and talking with people, these were some of the issues that were up at that time.

KM: Right. So Baylor was one of the G-5 sequencing centers. How were you guys dealing with credit?

Interviewee: C. Thomas Caskey

Date, location, method: 02 March 2012, Durham, NC, by phone

Interviewers: Kathryn Maxson, Robert Cook-Deegan

TCaskey: I think it still continues to be a problem for young people who have become part of these systems. How do you get distinguished? And how are you recognized? So I think it is a challenge, and of course there are a small number of leaders whose names come up in the press constantly and they receive the recognition. But many of the people who did a lot of the real work of the projects probably didn't get the kind of recognition you would have gotten if you were operating your own independent research program in an RO1 activity. I mean, the basic problem here is going industrial with an NIH project. That was a new challenge. But they were being pushed by an industrialist. If you wanted to beat Craig, then you better change your policies. You can tell me if I'm remembering these things incorrectly or not, but was the big picture whether NIH would be a player in this or is this going to be done by Craig?

BCD: And how are you viewing that, because you kind of had your experience at Merck and you also had your experience as the director of a genome center at Baylor that emerged as one of the G-5. How do you think about this?

TCaskey: I was absolutely confident that Craig had it right. Now, I'm going to divert from that. If you're going to do whole genome I think he had the right strategy of shotgunning and let the computers do the assembly. And of course that's proven to be correct. But I was confident he had that right from the start. It took NIH some time to realize that the computers were good enough, that the read lengths were adequate to be able to link things that had repeats and to be able to build the contigs. It took them some time to do it. And of course the community wanted to have this thing highly distributed so the money went into many different institutions. And that's a little bit of a political science funding issue, which was not an issue for Craig because it was a single site, a single corporation.

Now, I had...and this is where I was, I guess, I was urging doing something that did not win favor, and it came out of the EST experience. We were making progress in assembling the messenger RNA sequences, and I thought—this is where we ought to be working. It's 100-fold less complex; it doesn't have the damned repeats in it. We should be doing the expressed sequenced genome. That should be objective one. And when we get more efficient then we can take on the genome. But I didn't win on that one. I could not sell that concept at all. We tried and it was just falling on deaf ears. The enthusiasm was to do the whole genome. And of course that was technically extremely difficult and a lot more costly, 100-fold more costly, and technically a lot more difficult. So I was in a camp that ultimately didn't win the day on the first priority, which was, get the genes.

Now, to give credit where credit's due on the EST project, the EST sequences enabled the people who were doing the genome sequence to organize the genes into exons and introns. I mean, it was absolutely beautiful how that data, that sequence data, enabled you to take a genome sequence and to be able to figure out things very quickly. So without the EST database I think we'd have been damned

Interviewee: C. Thomas Caskey

Date, location, method: 02 March 2012, Durham, NC, by phone

Interviewers: Kathryn Maxson, Robert Cook-Deegan

slow in figuring out the genome. ESTs were extremely critical in understanding the organization of a genome and genes. If I had been king I probably would have sequenced all the genes and then come back and done the genome later. But the way it worked out, it worked fine.

KM: So what was 'agreed to' at Bermuda in 1996 was initially reported on the web by HUGO. All human genomic sequence information generated by centers funded for large-scale human genome sequencing should be freely available and in the public domain in order to encourage research and development and to maximize its benefit to society. Sequence assemblies should be released as soon as possible. In some centers assemblies of greater than 1kb would be released automatically on a daily basis. Finished annotated sequence should be submitted immediately to the public databases. Did you disagree with that?

TCaskey: My recollection of it was that there was a modification of that view. In other words, the centers had an amount of time to do the assembly and validate. And I had remembered six months, but maybe I'm incorrect on that. There was some wiggle room on that. I mean, the way you worded it, which I guess is the wording of the agreement, you would have been spitting out sequence information almost off the machines.

KM: Exactly, and you know, the NIH policy was six months. *[KM: The predecessor policy from the NIH, formulated in 1992 for mapping information, gave investigators 6 months to release their data.]*

TCaskey: Yeah, that is my recollection of what actually happened. Six months would give time for investigators to determine whether they were onto an original discovery. In other words, a gene and a function.

KM: So you were thinking that...so that was the policy from NIH, but that was from 1992. And so this in Bermuda was much different than what the NIH had been advocating previously. And it's your recollection that what was actually happening was that people were taking some time to make sure that the data weren't total crud and that the machines were working properly and that sort of thing. And there was indeed a lag time that was honored by the big centers.

TCaskey: That was my understanding of what took place operationally. Now you'd have to ask people like Richard Gibbs and others at the centers really how they were behaving. But my recollection was they had a window of about six months in which they could work with the assembly of the data. I mean, 1,000 base pairs, hell, that was a couple of runs on the machines. I can't imagine anybody releasing 1,000 base pairs as a unit. That's virtually nothing. You know, ask Richard about that. He would know what they did operationally.

Interviewee: C. Thomas Caskey

Date, location, method: 02 March 2012, Durham, NC, by phone

Interviewers: Kathryn Maxson, Robert Cook-Deegan

KM: Right, right. So I wonder, in light of your recollections of how this actually took place on the ground, what you feel the effects that this Bermuda agreement were on the conduct of science, particularly during the human genome project?

TCaskey: It ended up being a very wise decision and very successful, and here are the reasons that I think that. When this project was started it was incredibly opposed by cell biologists. And by the time the project got underway, virtually every cell biologist I know started the morning by going in, turning on the computer and looking at the latest entries because they were beginning to design their functional studies based upon the sequence information that was coming in. They could set up constructs; they could set up cell-based assays. They could really take a sequence and convert it into functional knowledge. So it enabled an entire research community to be able to be work on function, and this was a group of people who were skilled in different areas other than the informatics and the sequencing, but had tremendous contribution to what the genes do. So I think it was an incredibly successful decision.

KM: Right. You know, that's really interesting. [BCD], what do you think about that? That's really something we haven't heard from anyone else before.

BCD: I'm trying not to say too much because I'm in a very noisy room and I keep hitting my mute button.

KM: Yeah, no, no worries, I know that's what you were doing.

TCaskey: All you had to do was go to the departments of cell biology and ask the question, how important is the genome sequence information to your hypothesis-driven research? I can tell you, it's a major factor.

KM: And [BCD] already talked about this a little bit, but we've really been hearing a set of several really pretty consistent views, and one is that these Bermuda Principles, in quotation marks, were agreed to as a way to divvy up the work and divvy up what was going to be done by the different labs. And one was to be able to keep tabs on what people were doing, what people were outputting and whether or not they contributed what they said they were going to contribute. And another was a very political reason, and that was this is just good politics, right, and it made this data seem universal, and in a sense because it was the human genome project, it was something that should be shared amongst the entire scientific community. And what you're saying makes it sound like this was really worthwhile in a way that was outside even of the genomics community.

TCaskey: Remember, when we talk about the genome project, the human genome project is one, but the technology enabled any group to go out and do *E. coli*, yeast, nematode, *Drosophila*, and it just opened up the field of hypothesis-driven

Interviewee: C. Thomas Caskey

Date, location, method: 02 March 2012, Durham, NC, by phone

Interviewers: Kathryn Maxson, Robert Cook-Deegan

research to those cell biologists, cell and molecular biologists. I think it was incredibly stimulating to the field.

KM: Right. So at Baylor, I'm curious, did you ever have, did it seem like anyone was ever checking to make sure you were releasing your data on time? Was there any enforcement of the Bermuda Principles?

TCaskey: That's something you'd have to ask Richard about, because he was in on the calls. I think they were on almost weekly calls on productivity. It was horse race. You have to understand that. It became a horse race between the two programs, between Craig and the NIH program. NIH/Department of Energy. So they moved to, I think, weekly reportings on the progress and the output on each of the labs. But Richard would know that better than I would.

KM: You mentioned the impact of these principles on the research of cell biologists at the time. Along that idea line, do you think that this model for doing science amongst the human genome project folks has spilled over or affected the way that science is done? Not so much in the use of the data by the cell biologists, but I mean more in terms of an ethos of conducting science in an open manner. Do you feel like this has affected other areas of scientific research at all?

TCaskey: Well, I'm not sure. You've got to realize, I've come from the old school. And the old school is, you're an individual investigator, you have a creative idea, you go out, you do experiments and you create results as an individual. So the more modern-trained scientists operate in these pools of collaborating on data, the more they will have a different attitude toward it. I look back on people, though, that have gotten the Nobel and people who have gotten the Lasker and people who have gotten significant recognitions, and generally it's due to their intellect and their individual activity that pushes the field forward. Occasionally, you will have situations in which, like the sequence knowledge, the pooling of the data enables many investigators to work. But I'm still a person who likes to see an individual come up with a specific idea, test it and prove it. That's just old-fashioned; I'm just old-fashioned. But remember, I'm a physician too. We were trained to be individuals and to behave as an individual scientist as we cared for patients and did research. Not a lot of kumbaya in the days of my training. We were responsible for the activities.

BCD: So one of the other things that we make sure that we cover, and I'm reminded of this because it was your tape I had of the June 1986 Cold Spring Harbor meeting, although apparently you didn't remember that. But, it was your tape that was used as the transcript when I was writing *The Gene Wars*. Do you have any documents or do any people come to mind that we should talk to that we might not have thought of by only looking at the people who actually attended the meeting?

Interviewee: C. Thomas Caskey

Date, location, method: 02 March 2012, Durham, NC, by phone

Interviewers: Kathryn Maxson, Robert Cook-Deegan

TCaskey: Oh, Lord, I don't think I can help you on that. Let me tell you, I am terrible at retention of information because of what I did with my career. When I left Baylor College of Medicine I could not take records with me. They were the possession of the medical college when I went to Merck. When I left Merck the information was the possession of Merck and I could not take that information away. And the storage of all of my lab books and the schedules and agendas and so forth, they were warehoused at Baylor and I've never gone back to try to retrieve anything. They did it on some of the patents. They did go back and retrieve notebooks on some of the patents. But I'm just not a collector of that kind of information for technical reasons, and also it's not something that interests me very much.

KM: Right. Well we know we should talk to Richard Gibbs.

TCaskey: I'm just trying to think who...I'll tell you somebody who would be good to talk with, because he's in a different system other than human. I would talk to Hugo Bellen in *Drosophila*. He's also at Baylor, but he appreciated extremely early on the value of the sequence information to understand the functionality of *Drosophila*. And then of course they have these marvelous genome-wide insertional strategies that enabled them to interrupt and tag every gene. So he can tell you. He can tell you how much it accelerated *Drosophila* research. Without the sequence on *Drosophila* that field would have been old fashioned.

KM: Great. Well, [TCaskey], we've covered pretty much everything that's on our list, except for, were you by any chance at any of the later discussions about extending the Bermuda Principles to other areas of genomics and also other areas of biology, either in Fort Lauderdale in 2003 or Toronto in 2009?

TCaskey: No, but I've been in numerous conversations related to the patent issue on genes, disease genes. I can't tell you how many meetings I've attended on that. And that one's going to play out in the Supreme Court pretty shortly.

KM: Yeah. So, thank you so much. [BCD], unless I am missing anything ...

BCD: I had actually one question for you, which is, do you know how we can track down Alan Williamson?

TCaskey: I can get you an address for Al. Let me do that. I've kept up with almost all my old post-docs and pre-docs. I can get you an address for him.

BCD: That would be really helpful because he's kind of dropped off our radar.

TCaskey: Okay, I can find that. Okay.

BCD: All right, thank you so much.

Interviewee: C. Thomas Caskey

Date, location, method: 02 March 2012, Durham, NC, by phone

Interviewers: Kathryn Maxson, Robert Cook-Deegan

KM: Well great. Thank you so much.

TCaskey: Bye-bye.

BCD: Okay.

KM: Bye, take care, bye.

END OF RECORDING